Forensic Science International 272 (2017) 127–141

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

The identification of markers for Geoforensic HPLC profiling at close proximity sites

G. McCulloch[a,b,*], L.A. Dawson[c], M.J. Brewer[d], R.M. Morgan[a,b]

[a] UCL Security and Crime Science, 35 Tavistock Square, London WC1H 9EZ, United Kingdom
[b] UCL Centre for the Forensic Sciences, 35 Tavistock Square, London WC1H 9EZ, United Kingdom
[c] James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, United Kingdom
[d] BioSS, Craigiebuckler, Aberdeen AB15 8QH, United Kingdom

ARTICLE INFO

Article history:
Received 11 June 2016
Received in revised form 24 December 2016
Accepted 9 January 2017
Available online 17 January 2017

Keywords:
HPLC
Forensic geoscience
CDFA
Soil

ABSTRACT

Soil is a highly transferable source of trace physical material that is both persistent in the environment and varied in composition. This inherent variability can provide useful information to determine the geographical origin of a questioned sample or when comparing and excluding samples, since the composition of soil is dependent on geographical factors such as climate, bedrock geology and land use. Previous studies have limited forensic relevance due to the requirement for large sample amounts and unrealistic differences between the land use and geographical location of the sample sites. In addition the philosophical differences between the disciplines of earth sciences, for which most analytical techniques have been designed, and forensic sciences, particularly with regard to sample preparation and data interpretation have not been fully considered. This study presents an enhanced technique for the analysis of organic components of geoforensic samples by improving the sample preparation and data analysis strategies used in previous research into the analysis of soil samples by high performance liquid chromatography (HPLC). This study provides two alternative sets of marker peaks to generate HPLC profiles which allow both easy visual comparison of samples and the correct assignment of 100% of the samples to their location of origin when discriminating between locations of interest in multivariate statistical analyses. This technique thereby offers an independent form of analysis that is complementary to inorganic geoforensic techniques and offers an easily accessible method for discriminating between close proximity forensically relevant locations.

1. Introduction

Forensic geoscience is the scientific discipline that applies the techniques developed to study earth materials pertaining to the law and has applications in any legal context where earth materials may be able to help investigators, judges or jurors establish “what happened, where and when it occurred and how and why it took place” [1]. Since earth materials are highly transferable, persistent and present at a wide variety of crime scenes, forensic geoscience can be used in many scenarios to aid crime reconstruction, corroborate witness statements or verify suspect alibis [2].

Trace geoforensic materials recovered from a suspect, victim or crime scene can be analysed and interpreted in order to establish whether it is possible to discriminate between items or locations of forensic interest. Usually, the techniques used to do this are well established methods, developed in the earth science disciplines for the purpose of studying the geographical and geological phenomena, which have been retrospectively adapted to forensic work and typically test the physico-chemical characteristics of the inorganic fraction of geoforensic samples [3–5]. There are clear differences, both from a conceptual and pragmatic point of view, for instance in the sample size and the level of spatial and temporal precision required, between the problems and questions encountered in forensic casework and those encountered in earth science research, therefore careful consideration must be given to these philosophical differences in order to properly interpret the data generated by these techniques. Therefore, there is significant value in the development of analytical methods that incorporate the specific requirements of forensic casework [6–11].

In order to provide maximum weight to the conclusions drawn from a piece of physical evidence, it is recommended that the

[*] Corresponding author at: UCL Security and Crime Science, 35 Tavistock Square, London WC1H 9EZ, United Kingdom.
E-mail address: g.mcculloch.11@ucl.ac.uk (G. McCulloch).

http://dx.doi.org/10.1016/j.forsciint.2017.01.009

analyses performed test independent sample characteristics and currently it is the inorganic fragments which are predominantly utilised in geoforensic analyses. The uppermost layers of soils are rich in organic matter [12,13], which is comprised of living organisms, their intact remains and the organic compounds produced by their decomposition plus any synthetic organic compounds added to the soil [14,15] therefore there is a need for the greater use of techniques capable of the analysis of soil components other than the inorganic minerals [15,6,3,16]. There are a number of techniques designed to analyse organic compounds that are primarily concerned with their separation, identification and quantification, In this study, high performance liquid chromatography (HPLC) was chosen as a suitable analytical technique to assess the degree to which it could offer an additional approach to the characterisation and discrimination of forensic soil samples. Not only is HPLC equipment, and the expertise to run this type of analysis, readily available in commercial forensic and analytical laboratories, but in addition, initial studies have shown potential for its application in forensic soil analysis [16,17–19].

Recent research into the feasibility of using HPLC for geo-forensic analysis [20] has addressed some of the practical and philosophical issues identified in the previous research, by reducing the sample size required for analysis to 250 mg, and simplifying the analytical procedure to reduce sample preparation run times. In addition, the HPLC technique presented by McCulloch et al. [20] demonstrated that excellent discrimination was feasible between samples derived from forensically relevant close-proximity locations, such as the entrance and exit locations of a defined crime scene and possible alibi sites. This newly developed method was found to be appropriate for comparing trace soil samples for

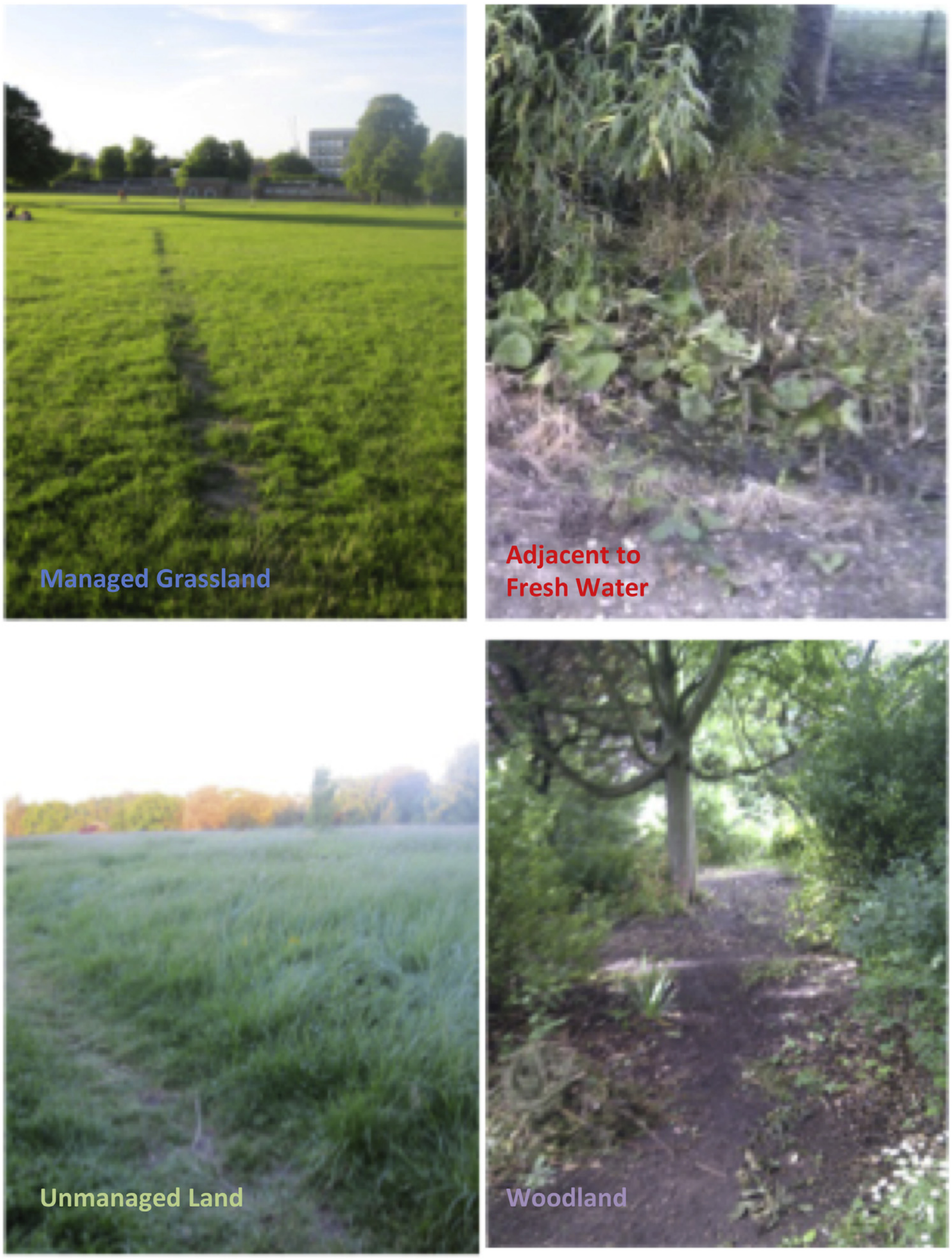

Fig. 1. Sampling locations within Brockwell Park, London.
Photographs taken at each of the four sampling locations at the Brockwell Park site in London.

the purposes of excluding a crime scene, an alibi site and unknown samples for application in criminal cases. These contextual details are important, since they affect the considerations required for the appropriate interpretation of the evidence. A great many analytes were detected in the soils used in the McCulloch et al. study [20], which poses a significant barrier to implementation of the technique in routine casework as the required data analysis was too labour-intensive to be considered practical for routine analyses.

The aims of this study were to test a new analytical method and determine whether the underlying geology of a sample site affects the ability of the method to discriminate samples of interest. In addition, this study aimed to select a significantly reduced number of useful target analytes for multivariate statistical analysis, in order to provide the same excellent discrimination between sites offered by the existing method [20], but with a more easily implementable data analysis method, thereby increasing the potential use of HPLC as a profiling tool for geoforensic samples in casework.

2. Methodology

2.1. Site description

All the sites chosen were well-established municipal parkland, intended and maintained for public recreational use by the local authority, and each site was located on contrasting underlying

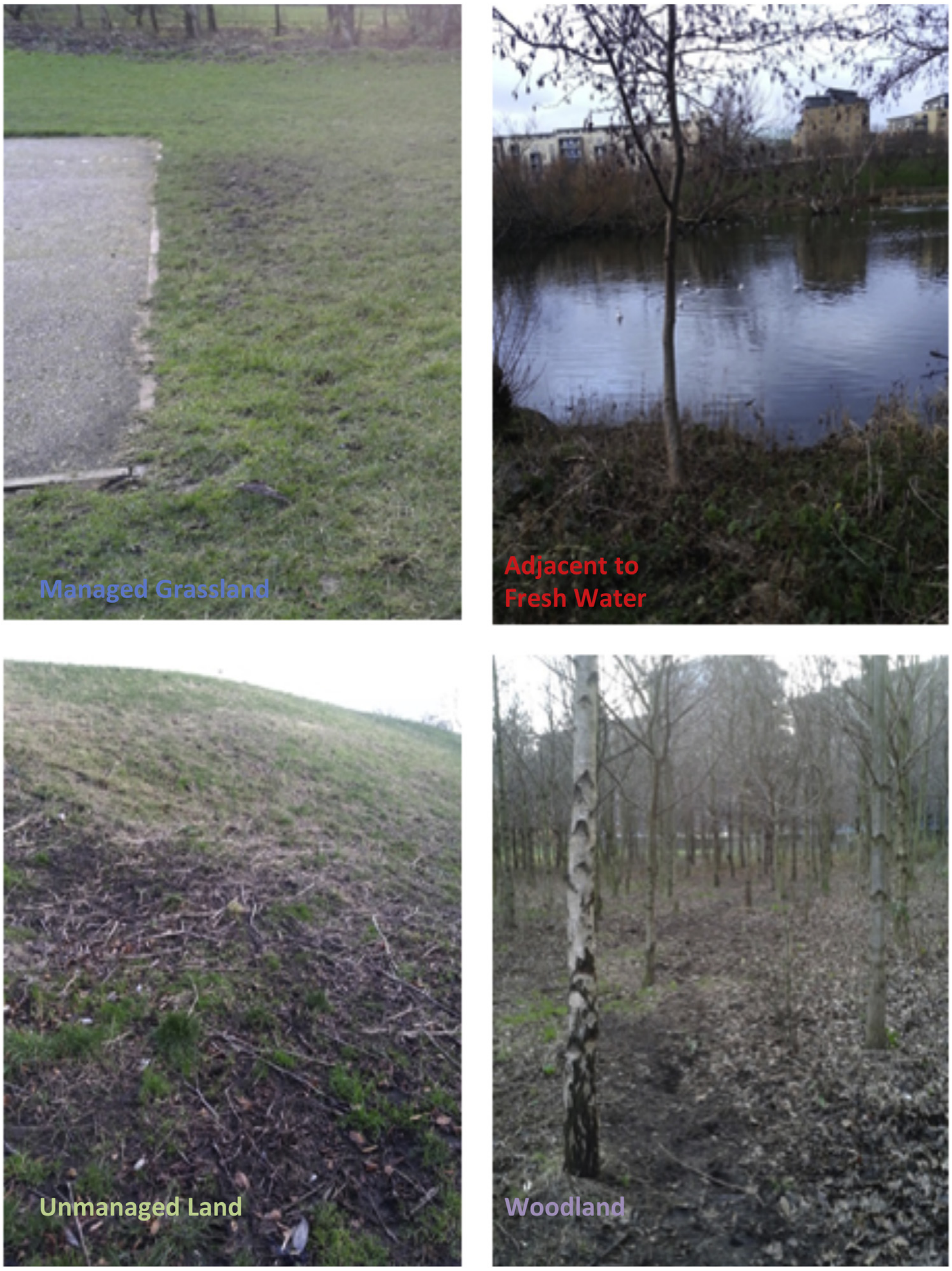

Fig. 2. Sampling locations within Lochend Park, Edinburgh.
Photographs taken at each of the four sampling locations at the Lochend Park site in Edinburgh.

Table 1
Visible and land use characteristics of sample locations.

Location (land use type)	Description
1: Managed grassland	A flat area of well-maintained, cut grass used for exercise and sporting activities.
2: Adjacent to fresh water	A flat area of miscellaneous wild vegetation, immediately adjacent to a fresh-water pond, housing various water fowl, and with restricted pedestrian access.
3: Unmanaged land	A natural footpath through a sloping area of miscellaneous wild vegetation, such as wild flowers and grasses.
4: Woodland	A natural footpath through a flat area of relatively bare earth with a dense canopy of primarily deciduous trees, shrubs and localised leaf litter, immediately adjacent to a residential area and used as a thoroughfare to the park entrance.

geology. Three sites in the UK were selected, Brockwell Park in London (Fig. 1), Lochend Park in Edinburgh (Fig. 2), Craigiebuckler Estate in Aberdeen (Fig. 3), while one site was located in the USA, in Central Park, New York City (Fig. 4). At each site, four close proximity, but distinct sampling locations, were chosen that represented both potential alibi sites and potential crime scenes. The locations were recreational areas where a person could legitimately come into contact with earth materials, or secluded spaces and thoroughfares, which could provide opportunities for crimes to be committed and were selected utilising case work

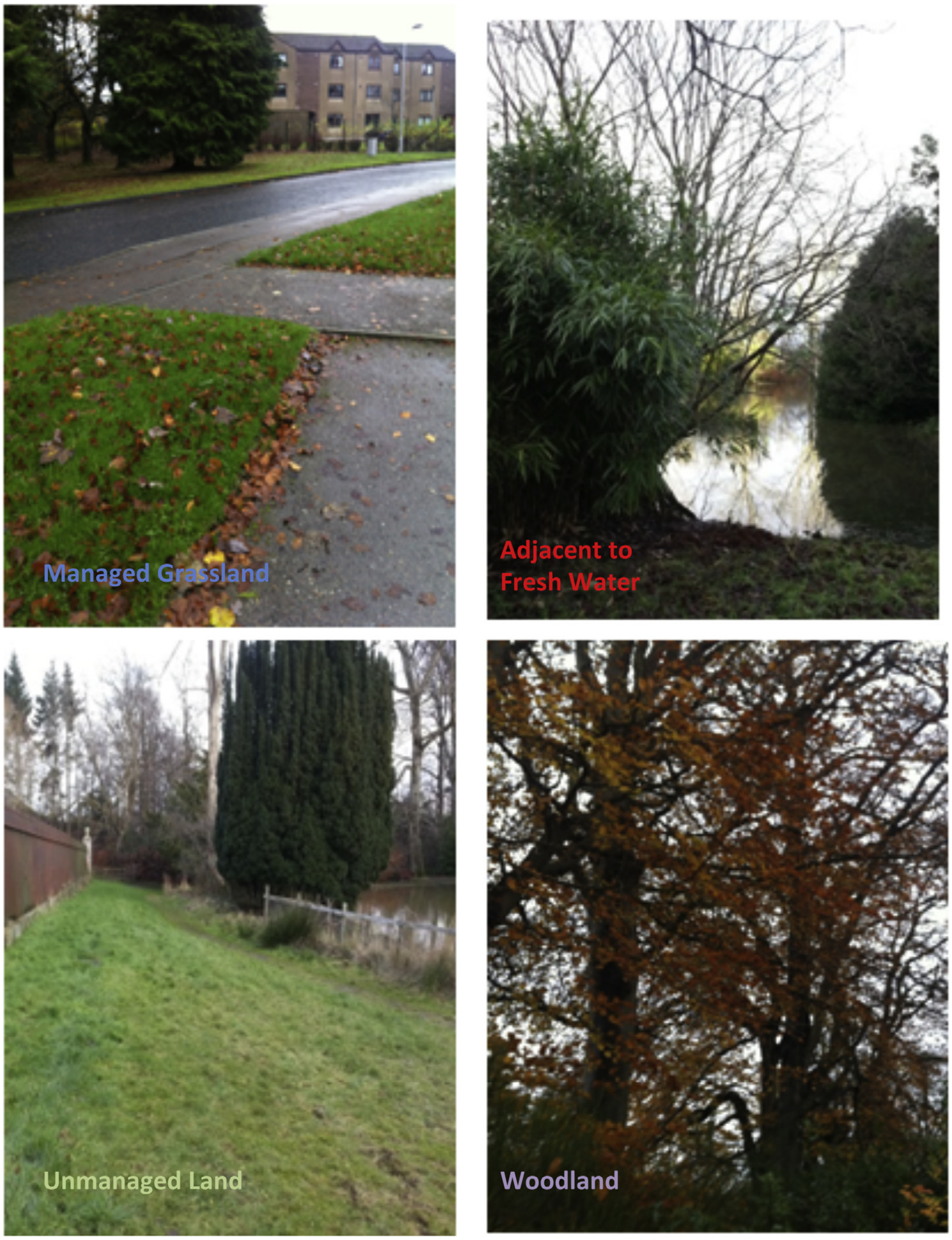

Fig. 3. Sampling locations within Craigiebuckler Estate, Aberdeen.
Photographs taken at each of the four sampling locations at the Craigiebuckler site in Aberdeen.

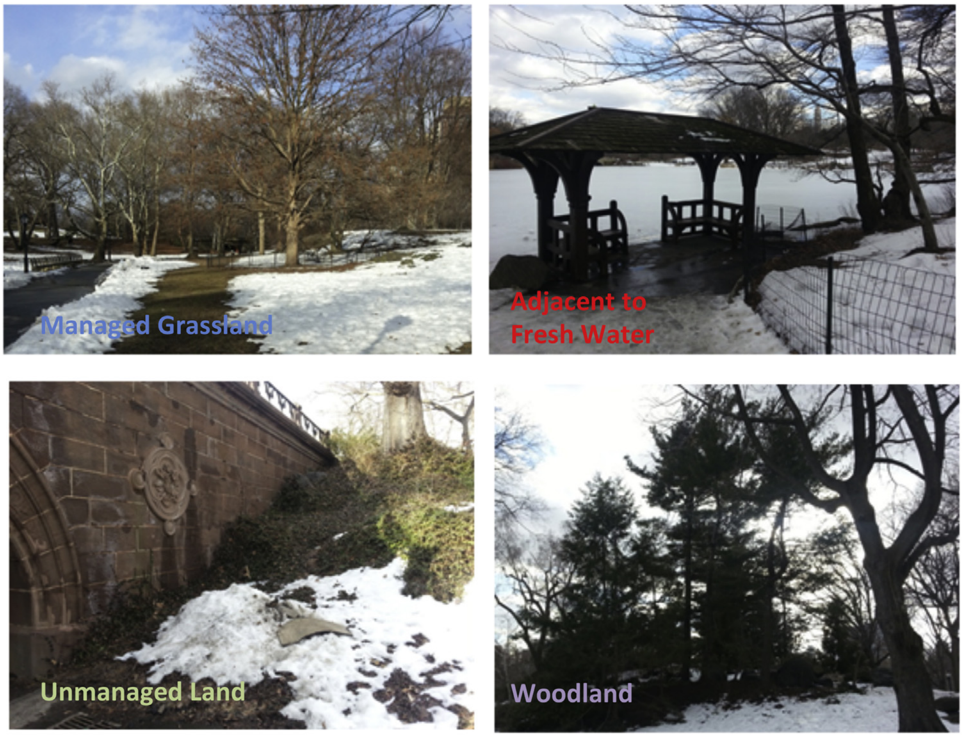

Fig. 4. Sampling locations within Central Park, New York City.
Photographs taken at each of the four sampling locations at the Central Park site in New York City.

experience. The descriptors detailed in Table 1 were applicable to each location, at each site.

2.2. Sample collection and preparation

Five samples were collected from each location in order to assess intra-location variability, using the grid pattern suggested for sampling footprints and tyre tracks by Pye [21]. In accordance with Simmons [22], samples were gathered using a stainless steel spatula, removing any surface turf or gravel, where present. Approximately 5 g of in situ topsoil (c. 0–1 cm depth) was collected at the corners and central point of a 1 × 1 m square grid. All samples were stored in breathable containers and allowed to air dry prior to use.

250 mg of dry soil was added to a 1.5 ml sterile, DNA free, polypropylene centrifuge tube and 0.5 ml gradient grade acetonitrile was added by pipette. The tubes were placed in a sonic bath for 20 min then centrifuged for 15 min at 13,000 rpm. The supernatant was then passed through a 0.22 μm PTFE syringe filter into an HPLC vial.

2.3. Instrument parameters

Samples were injected onto an Agilent 1100 HPLC system with DAD detector, using UHQ water as mobile phase A and gradient grade acetonitrile as mobile phase B, which had been degassed by sonication prior to use. A series of method development experiments were first performed on the HPLC method used in previous studies [20] to yield useful, discriminatory profiles. From these experiments, the optimum column, mobile phase and gradient parameters required to maximise the number of peaks detected per run and to reduce the overall sample analysis time were determined. Table 2 details the instrument parameters selected for use after method development.

2.4. Data analysis

2.4.1. HPLC Data analysis

The HPLC profiles obtained from each sample were examined using Agilent Chemstation software (version B.04.01), and were found to contain hundreds of individual, closely eluting peaks, many of which were close to the limit of quantification (LOQ). In order to standardise the peak heights, the data were first scaled to correct for differences in sample concentration and the data adjusted to the theoretical response for a 500 mg/ml solution, in addition the peaks below LOQ were removed from the data set. In order to reduce the number of variables and simplify the data analysis, and to minimise error rates, two subsets of peaks were subsequently chosen from this data set to use as markers, these are identified by their retention time in Table 3.

The first set (set A) contained the 20 largest peaks observed in the data, since these were the clearest peaks to identify and

Table 2
HPLC parameters.

Injection volume	50 μl		
Column	Waters Xbridge C18, 3.5 μm, 150 × 4.6 mm at 30 °C		
Gradient	Time (min)	% Mobile phase A	% Mobile phase B
	0.0	53	47
	3.0	45	55
	24.0	26	74
	29.0	2	98
	31.0	2	98
	32.0	53	47
	35.0	53	47
Flow rate	1 ml/min		
Detector settings	254nm, bandwidth 4 nm, peak width >0.1 min		

Table 3
Retention times of marker peaks.

Marker set	Peak retention times (minutes)
A	4.4, 9.0, 9.4, 10.0, 10.8, 11.6, 12.2, 12.6, 13.6, 14.2, 15.0, 15.5, 15.8, 18.8, 19.6, 20.3, 23.6, 24.3, 37.3, 30.4, 30.8
B	1.9, 4.4, 6.7, 12.2, 13.2, 13.7, 15.0, 19.1, 24.5, 26.9, 28.5

quantify. In addition, as they represent the major components of the sample, these peaks were considered to be the good analytical targets as they could potentially allow for a decrease in the working sample concentration, and therefore reduction of the sample size, in future studies.

The second set (set B) were selected using the R [23] "subselect" package [24] to analyse the data set and to identify subsets of a smaller number of variables which gave equally good classification accuracies as the full data set, based on the Wilks' lambda values. In multivariate analyses, Wilks' lambda is analogous to the F-value in univariate analysis of variance, and is the test statistic against which the significance of the differences in the group means can be assessed [25].

This process produced ten subsets of three or four peaks each, which, when used as variables in a leave-one-out classification, have near-perfect error rates. Amongst these ten peak sets, some of the peaks appeared more than once, and eleven distinct peaks in total were selected as useful markers during this data analysis step.

For each of the markers in these sets of peaks, the mean peak height of the five replicates analysed were calculated and plotted for samples from each of the four locations at each site, using Microsoft Excel, along with bars displaying the standard error in the means, in order that the regions of variability in the profiles could be more easily visualised.

2.4.2. Canonical discriminant function analysis (CDFA)

After integrating and processing the chromatographic data, CDFA was performed on the data from the five replicate samples at each of the locations, using SPSS to determine the accuracy and precision with which these markers allow samples to be grouped according to their location within each site. This data analysis

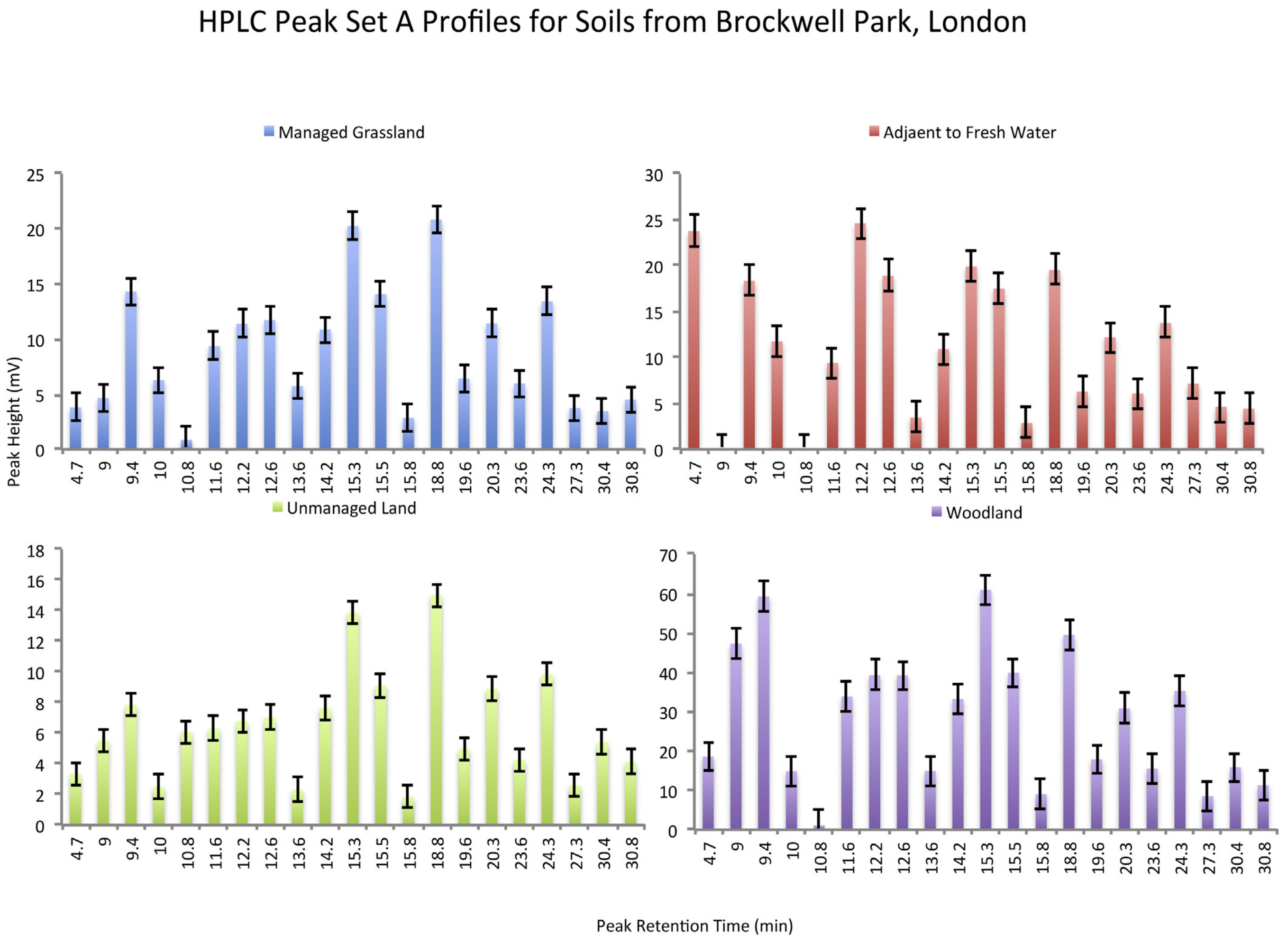

Fig. 5. HPLC profiles for Brockwell Park, London- peak set A.
Profiles of the mean (n = 5) height (bars display the standard error of the mean) for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the London samples using peak set A.

technique used each HPLC peak as a predictor variable, and each sample location as a grouping variable. The data for each sample were used by the software to generate functions, which are linear combinations of the variables that maximise the difference between each location. Each sample was then plotted using the scores for each function as co-ordinates, to create a scatter plot where samples of similar composition clustered closely together, allowing groups of samples, and the relative degree of difference between groups, to be visualised. The functions were then used to assign each sample in the data set to a particular location, based on the scores for each function, and the accuracy of classification was determined by comparing the predicted sample location to the true sample location.

3. Results

3.1. HPLC profiles for peak set A

All four locations within Brockwell Park, London could be distinguished by the profiles of HPLC peak set A (Fig. 5). The size order for the peaks at 4.7, 9.4 and 10 min were distinct for samples adjacent to fresh water, as was the absence of the peaks at 9 and 10.8 min. Managed grassland could be distinguished from woodland and unmanaged land by the size of the peak at 9.4 min, the ratio of this peak compared with the peak at 9 min was larger for managed grassland, at 3:1, than for both unmanaged land and woodland, at 1.3:1 and 1.4:1, respectively. The relatively high ratio of 2.5:1 for the peak at 15.3 min, relative to the peak at 9 min, for woodland samples separated these samples from unmanaged land where the ratio was 1.2:1.

It was possible to discriminate all four locations within Lochend Park, Edinburgh on the basis of the profiles of HPLC peak set A (Fig. 6). The presence of a peak at 10.8 min was a useful discriminator for the soils adjacent to fresh water and the soils from unmanaged land were the only samples that did not contain a peak at 15.8 min. At the managed grassland location, the ratio of the peak height at 9 min relative to the peak at 4.7 min was much larger, at 3.5:1, than for woodland samples, where this ratio was 1.5:1.

The profiles for HPLC peak set A (Fig. 7) allowed each of the four locations in Craigiebuckler Estate, Aberdeen to be distinguished visually. The largest peak present in the managed grassland samples was the peak at 9.4 min, which was distinct from the other three locations. The profiles of soils adjacent to fresh water differed from the other locations in that the largest peak was at 10.8 min while the large relative height of the peak at 30.8 min was distinctive of woodland soil profiles. The profile of the samples from unmanaged land was distinctive with the highest peak at 30.4 min, the presence of the peak at 9 min and the absence of a peak at 12.6 min.

In Central Park, New York City, all four locations could be discriminated on the basis of their HPLC profiles for peak set A

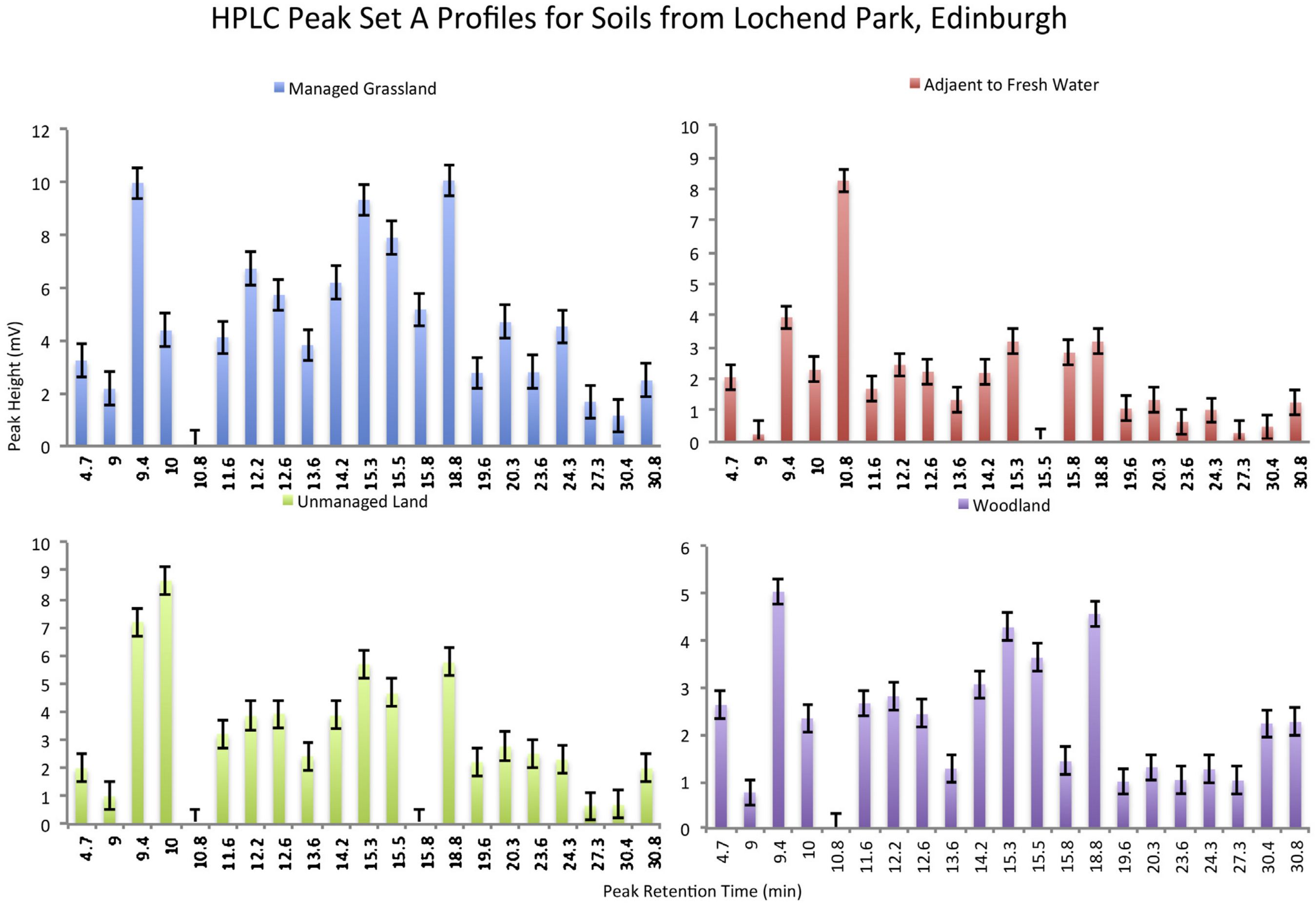

Fig. 6. HPLC profiles for Lochend Park, Edinburgh- peak set A.
Profiles of the mean (n = 5) height (bars display the standard error of the mean) for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the Edinburgh samples using peak set A.

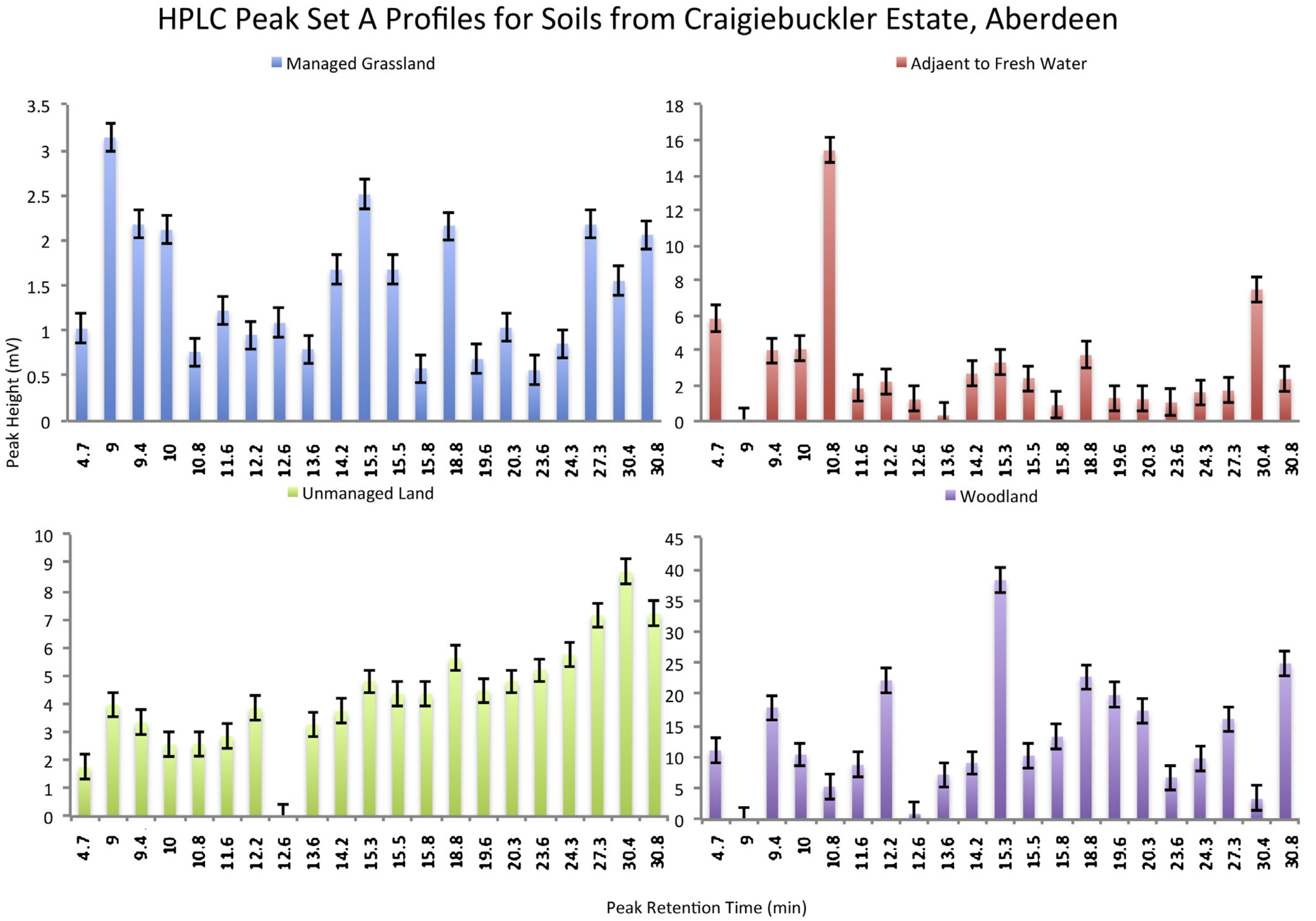

Fig. 7. HPLC profiles for Craigiebuckler Estate, Aberdeen- peak set A.
Profiles of the mean (n = 5) height (bars display the standard error of the mean) for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the Aberdeen samples using peak set A.

(Fig. 8). The location adjacent to fresh water could be distinguished from the other locations through the absence of a peak at 10.8 min, while the absence of the peak at 9 min was a unique feature of the soils from unmanaged land at this site. The peak at 9 min was larger than the peak at 4.7 min, at 9 mV and 3 mV, respectively, for the managed grassland, while for the woodland location the peak at 4.7 min was larger at 12 mV compared to 5 mV at 9 min. In addition, the peaks were generally three times larger for the woodland soils, ranging from 5 to 60 mV, than for managed grassland where the peaks ranged from 1 to 15 mV.

3.2. HPLC profiles for peak set B

The HPLC profiles for peak set B allowed all four samples locations within Brockwell Park, London to be discriminated visually (Fig. 9). The large size of the peak at 6.73 min relative to the peak at 12.2 min distinguishes managed grassland from all other sample locations, while the large size of the peaks at 1.9 min compared to all other peaks is distinctive of the profiles in soils adjacent to fresh water. The profiles of soils from unmanaged land and woodland were visually more similar, however on closer inspection the unmanaged land samples can be discriminated due to the presence of the peak at 24.5 min which is absent in woodland samples.

The profiles for HPLC peak set B were not as easily distinguishable at each of the four locations in Lochend Park, Edinburgh (Fig. 10). The woodland samples were distinctive in that they had large peaks at 1.9 min, which were approximately ten times the size of the next largest peaks at 12.2 and 19.1 min. The soils adjacent to fresh water were distinct with the two largest peaks at 1.9 and 19.1 min, which were similar in size. The largest peak in the managed grassland profiles was approximately twice the size of the same peak in the unmanaged land profiles, in addition, the small peaks present in the managed grassland profiles at 6.7 and 28.5 min were absent in the samples from unmanaged land.

The samples from the four locations within the Craigiebuckler Estate, Aberdeen were easily distinguished using peak set B (Fig. 11). Comparison of the retention time of and ratio between the two largest peaks at each location was useful in grouping the samples. For managed grassland the largest peak was at 1.9 min and next largest at 12.2 min, but soils adjacent to fresh water had two large peaks of similar size, at 1.9 and 28.5 min. The ratio of the largest peak at 12.2 min to the next largest peak at 1.9 min was greater for woodland soils, at 8:1, compared to 3:1 for the unmanaged location, and woodland soils were also missing the peaks at 4.35 and 6.73 min that were present at the unmanaged location.

The profiles obtained for peak set B also varied across the four locations in Central Park, New York City (Fig. 12). Soil profiles for unmanaged land could be separated from the other three locations by the absence of the peak at 1.9 min while samples from managed grassland could be distinguished from those for woodland by the size order of the peaks at 1.9, 4.35 and 6.73 min. The peak at

Table 4
Canonical discriminant function results.

HPLC profiles	Classification accuracy %	Wilks' lambda significance test of differences in group means			% Variance function 1	% Variance function 2	% Variance function 3
		1–3 p=	2–3 p=	3 p=			
London							
A	100.0	<0.001	.002	.034	89.7	7.0	3.3
B	90.0	<0.001	.041	.684	84.8	13.9	1.3
Edinburgh							
A	100.0	<0.001	<0.001	.022	88.9	8.3	2.7
B	100.0	<0.001	.018	.397	62.4	32.5	5.0
Aberdeen							
A	94.7	.001	.147	.531	90.6	7.8	1.5
B	100.0	<0.001	<0.001	.014	97.4	2.4	0.2
New York							
A	100.0	<0.001	<0.001	.005	92.3	5.8	1.9
B	100.0	<0.001	<0.001	.071	73.3	24.1	2.5

12.2 min was 1.87 mV and was 33% smaller than the 2.78 mV peak at 1.9 min for soils adjacent to fresh water, while it was far larger than the peak at 1.9 min for managed grassland with peak heights of 5.49 mV and 0.94 mV, and for woodland the peak heights were 10.3 mV and 4.19 mV, respectively.

3.3. CDFA results

Full details of the CDFA results are provided in Table 4 and the scatter plots produced for both peak sets are shown in Fig. 13.

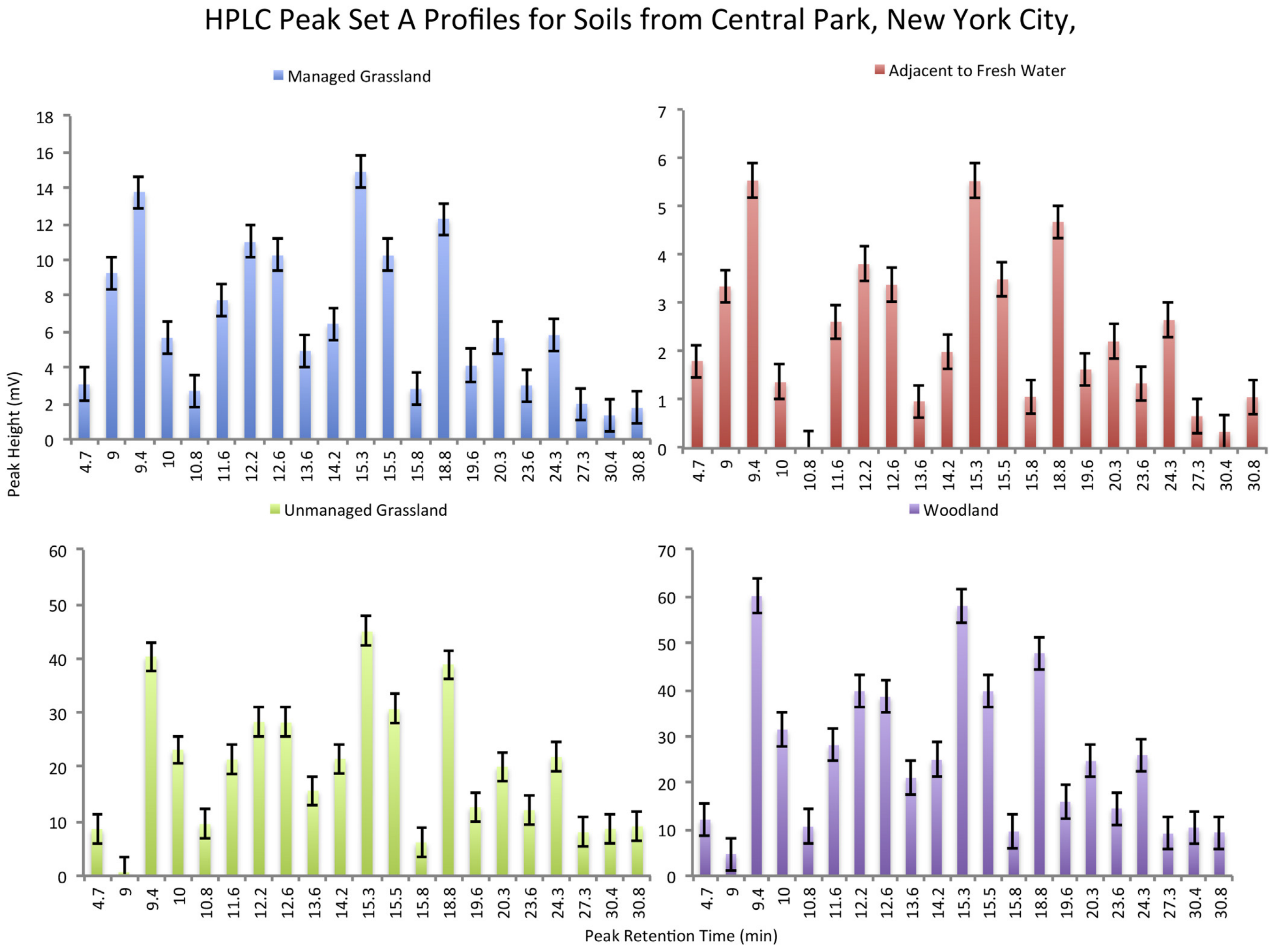

Fig. 8. HPLC profiles for Central Park, New York City- peak set A.
Profiles of the mean (n = 5) height (bars display the standard error of the mean) for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the New York City samples using peak set A.

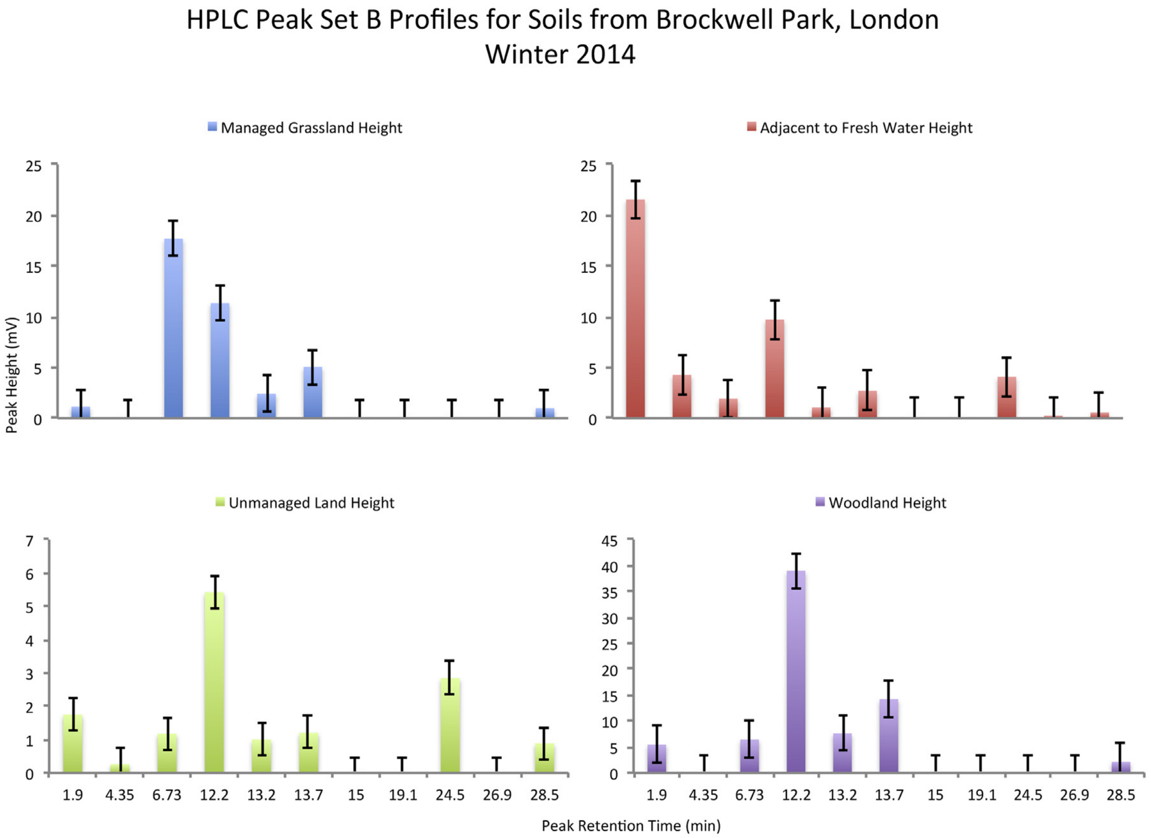

Fig. 9. HPLC profiles for Brockwell Park, London- peak set B.
Profiles of the mean (n = 5) height (bars display the standard error of the mean) for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the London samples using peak set B.

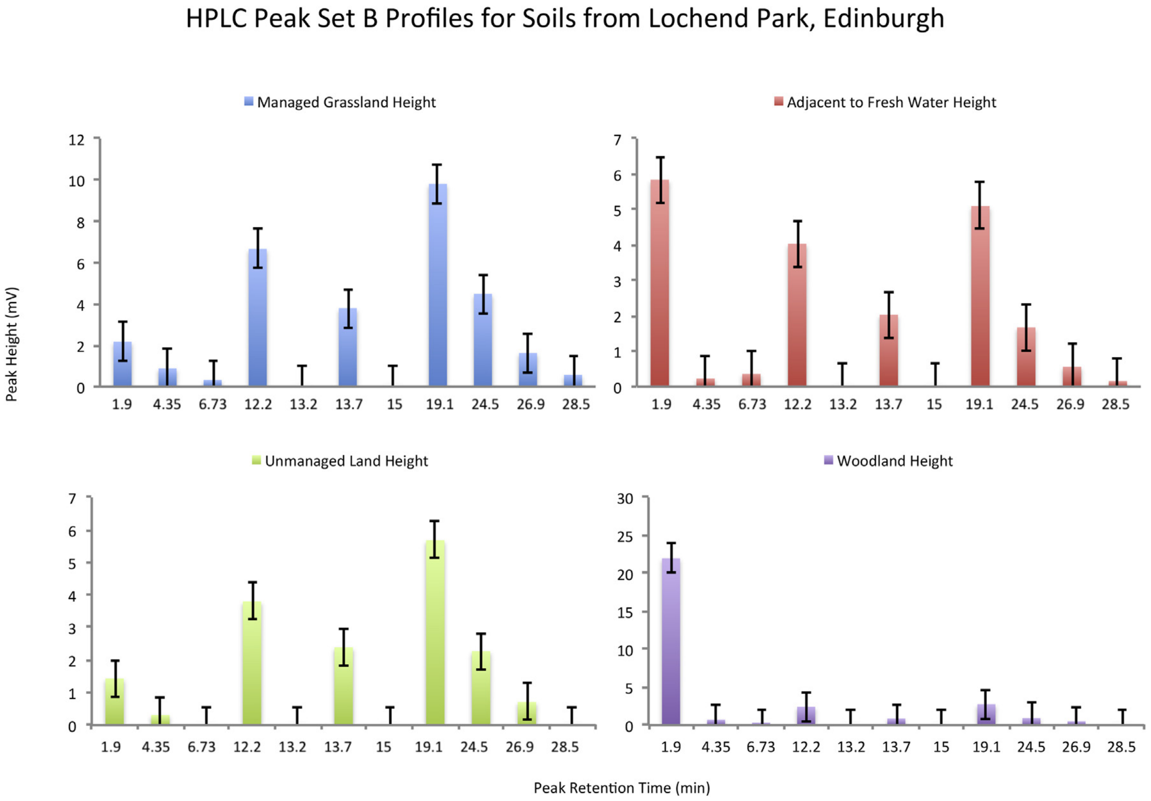

Fig. 10. HPLC profiles for Lochend Park, Edinburgh- peak set B.
Profiles of the mean (n = 5) height (bars display the standard error of the mean) for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the Edinburgh samples using peak set B.

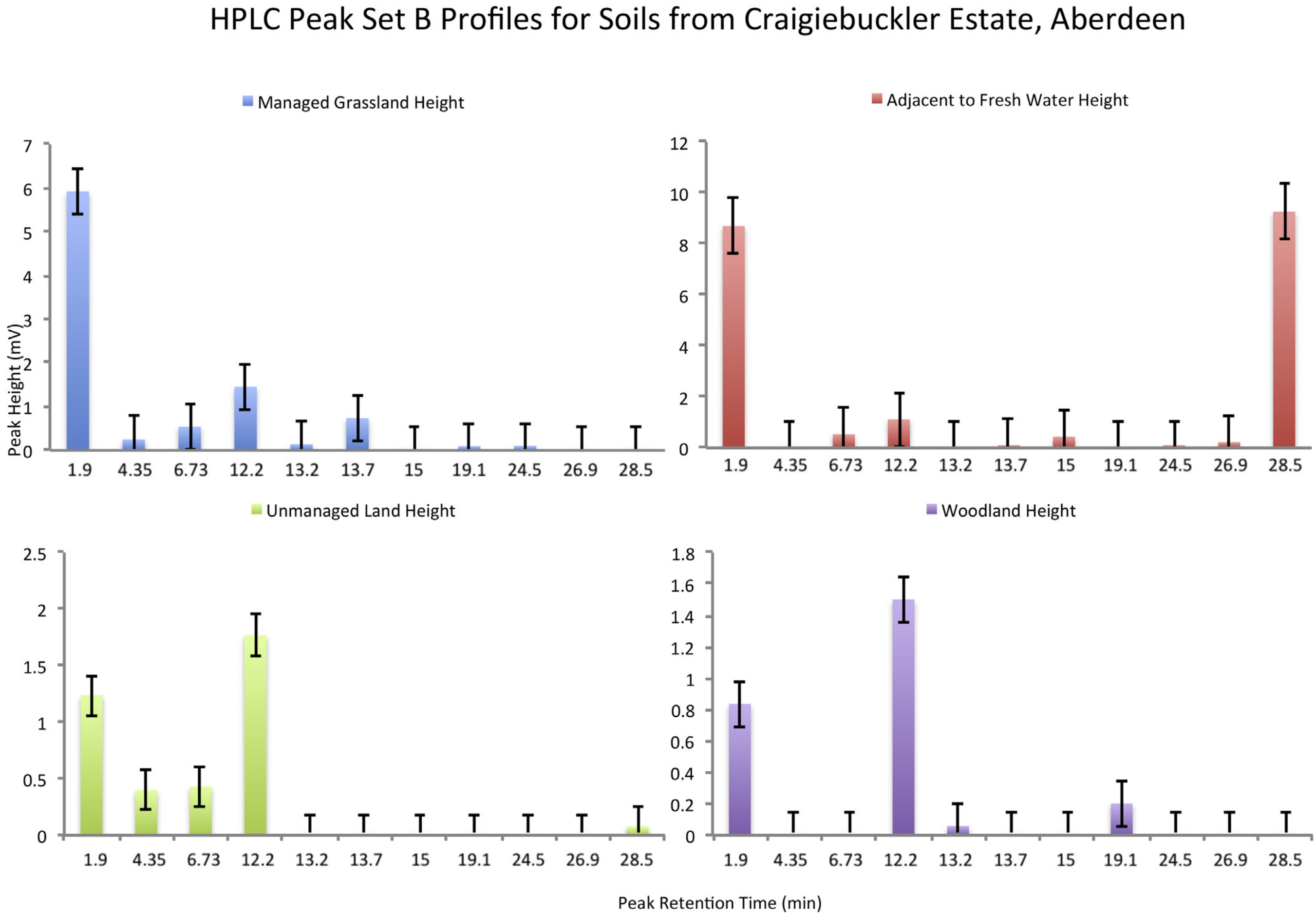

Fig. 11. HPLC profiles for Craigiebuckler Estate, Aberdeen- peak set B.
Profiles of the mean (n = 5) height (bars display the standard error of the mean) for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the Aberdeen samples using peak set B.

Using HPLC peak set A (Fig. 13), all but one sample was classified to the correct location across all four sites. One sample from the unmanaged location in Craigiebuckler Estate, Aberdeen was misclassified as having originated from the woodland location, giving an overall accuracy rate of 94.7% for this peak set at this site. The discriminant functions gave rise to sample groupings that were statistically significant ($p < 0.001$ for London, Edinburgh and New York City, $p = 0.001$ for Aberdeen).

For HPLC peak set B, all the samples were correctly classified at the Aberdeen, Edinburgh and New York sites. Two samples from Brockwell Park, London were incorrectly assigned to groups using the functions generated. One sample from managed grassland was predicted to belong to the unmanaged land group, while one sample from the location adjacent to fresh water was incorrectly assigned to the managed grassland soil group. This resulted in a grouping accuracy rate of 90.0% for the London site. As with peak set A, the discrimination of sample groups resulting from the functions produced in this analysis, was also statistically significant ($p < 0.001$).

4. Discussion

Peak sets A and B both provided ways to distinguish the four close-proximity locations at each site based on visual assessment of the profiles, by peak ratios or the presence or absence of a particular peak. These results were consistent across the range of geographical locations used in the study, which indicates that the technique has the potential to be applied across a wide range of geographical locations throughout the UK (and potentially internationally) as a complementary form of analysis to inorganic techniques for the discrimination of close proximity sample locations.

However, it was not possible to discern any commonalities between similar location types at the different sites, for instance woodland soils were not shown to share similar profiles across all four parks, nor were any specific land use markers identified, suggesting that at this stage the HPLC technique is not yet able to provide intelligence related to the land use of the provenance location of an unknown sample. Given these results it is therefore, unlikely that geoforensic evidence from a particular location type (e.g. unmanaged grassland) at another site would yield an HPLC profile analogous to that of soils with the same location type in this study, which demonstrates the need, at this time, for cautious interpretation of the results for these samples when using this method.

Nonetheless, with only small databases of 20 samples, from each of the 4 sites presented here, both HPLC peak sets A and B offered outstanding discrimination in the context of this study, where the task was to compare and exclude evidentiary samples from within one particular place of forensic interest. Measurement of these specific peaks not only provides an accurate and informative method of comparing organic soil characteristics at

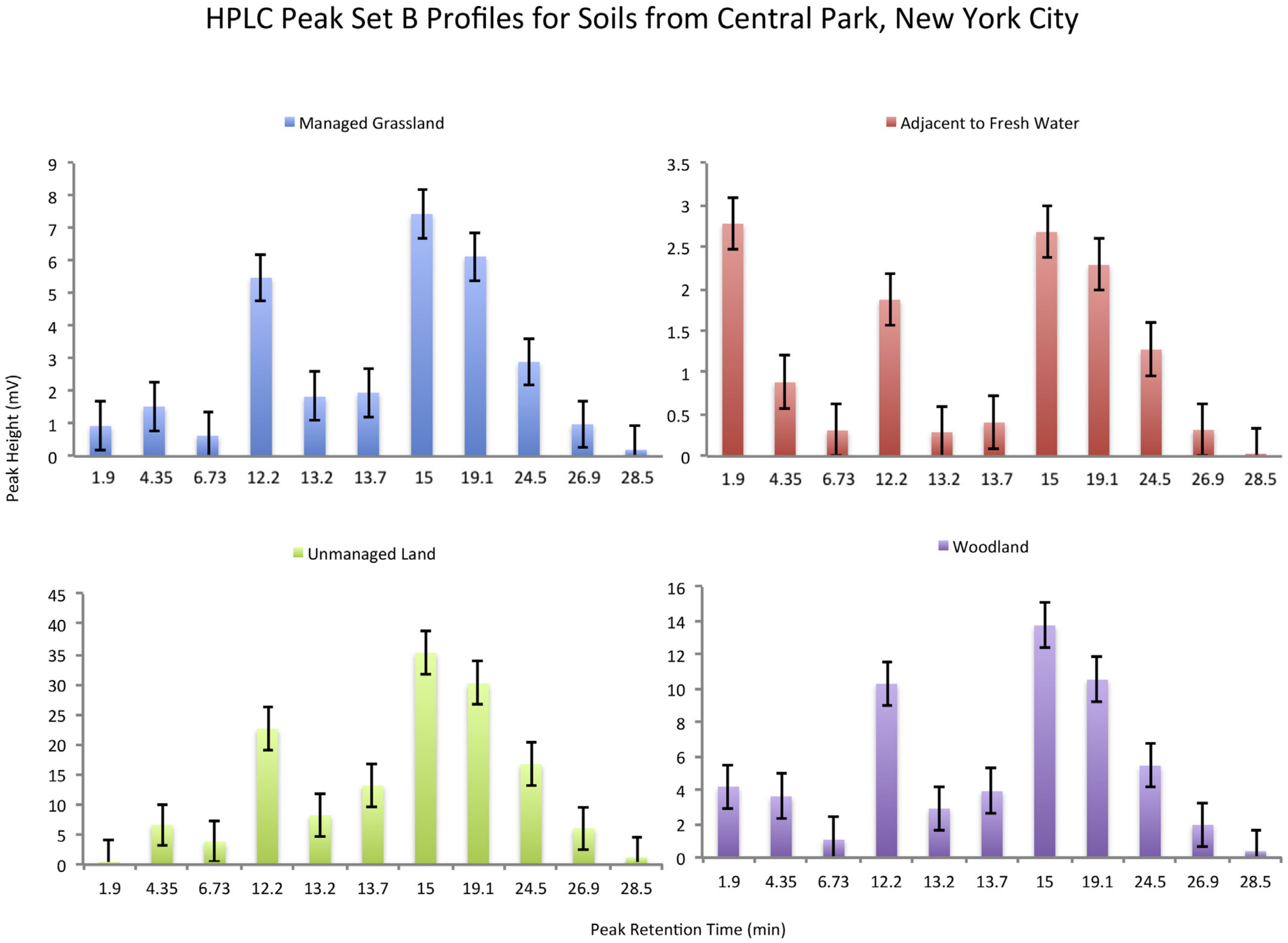

Fig. 12. HPLC profiles for Central Park, New York City- peak set B.
Profiles of the mean (n = 5) height (bars display the standard error of the mean)for 500 mg/ml solutions of dry soil in acetonitrile versus retention time, for the New York City samples using peak set B.

locations within a site, but when combined with further processing of the data using CDFA the ability of the technique to discriminate between these close proximity locations was excellent.

The new instrumental parameters used in this study reduced the sample analysis time by 30% compared to the method presented previously by McCulloch et al. [20] and the improved data analysis method outlined in this study significantly increases the potential impact of this HPLC method as a tool for the discrimination of geoforensic samples. Whilst similar levels of accuracy were achieved for the London site in a previous study [20], the data processing step was labor intensive, due to the large number of peaks observed in the raw chromatographic data, and each peak required manual integration and classification for every analysis [19]. This present study has shown that equally high grouping accuracy can be achieved by measuring only a smaller set of peaks, pre-selected for their discriminatory value, which consequently reduces the time taken to analyse the data from a number of weeks for the McCulloch et al. study [20], to less than one day. The confirmation provided here of the suitability of these two peak sets for use as markers for geoforensic profiling allows the use of automated integration and peak identification in future studies, which could allow the data to be prepared in seconds by the chromatography software ready for CDFA analysis in SPSS, potentially reducing the data analysis time to a few minutes.

The use of automated data processing on these selected peaks would also improve the precision between the replicates at each sample location, as the software ensures greater consistency in the integration of each peak. Furthermore, with fewer peaks to quantify per sample, it is easier to accurately assign the peaks of interest and reduce misclassification errors resulting from coeluting peaks and poor chromatographic resolution.

There was noticeable variability in the data, evident from the error bars displayed in Figs. 5–12, which was expected due to the natural variability in soil. It is likely that this variation would have been reduced had the five samples for each location been homogenized then sub-sampled, however since there was no pooling of the five replicates, in order to offer the most appropriate forensic context, the results are a better reflection of the true variability within each location. This approach is also applicable to when contact point sampling is required, for instance at a footmark. Homogenization of samples must only be performed after careful consideration of specific case circumstances [9], since there may be small quantities of diagnostic or characteristic compounds, that are essential to the interpretation of the results, present in a discrete soil aggregate or individual sample point, and this information may be lost if sample mixing dilutes such compounds to below the limit of detection for the method. Understanding the degree of variability in the profiles at a location

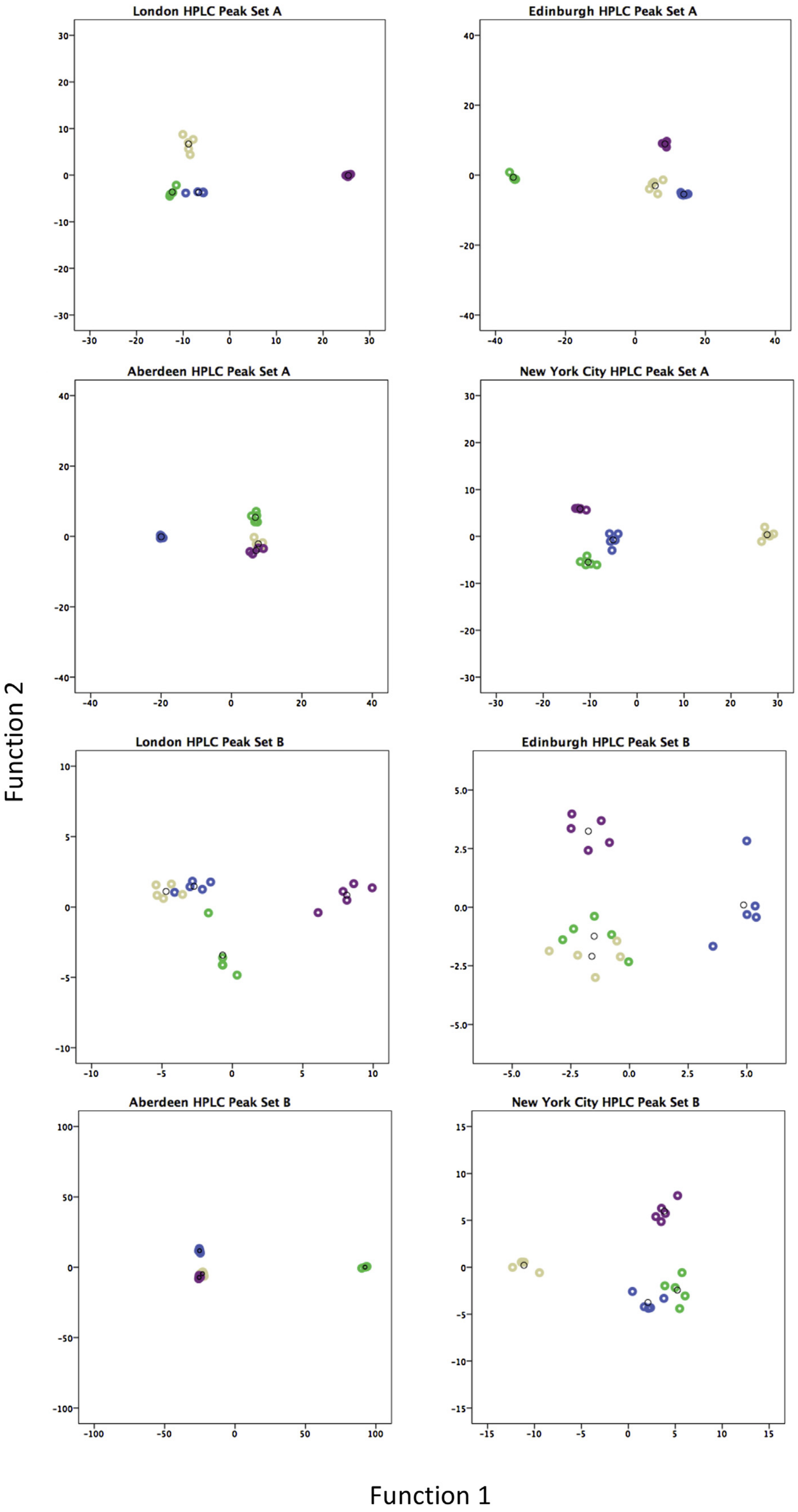

Fig. 13. Canonical discriminant function plots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.) Scatter plots showing the scores for each sample from managed grassland (blue), adjacent to fresh water (green), unmanaged land (yellow), and woodland (purple) and the centroids for each group (black) for the first two canonical functions.

of forensic interest is essential when making comparisons of control and questioned samples, therefore preservation of intra-location variation is key to appropriate interpretation of geo-forensic evidence.

The classification accuracy rate achieved for the two sets of marker peaks was very high for both sets of peaks, with both peak sets A and B correctly assigning 100% of the samples to their location of origin at three out of the four sites studied, at the London, Edinburgh and New York City sites for peak set A and the Aberdeen, Edinburgh and New York City sites for peak set B. The results were slightly improved for peak set A which misclassified only one sample across the whole study, compared to peak set B which misclassified two samples. In this regard, set A can be said to offer slightly superior results for this data set, however the data for peak set B had the advantage of being much more easily interpreted by visual examination of the profiles, due to having fewer variables to compare. With fewer peaks to identify, classify and analyse, set B also offers the added benefit of reduced data analysis time compared to set A.

In order to ultimately implement the HPLC technique in forensic casework in the UK, it will be necessary to validate the technique in accordance with the Forensic Science Regulator's guidelines, which would involve extensive investigation of both source and activity level propositions. It is first necessary to establish that the HPLC method is suitably accurate and precise to identify and quantify the marker peaks, by preparing isolating and purifying extracts of the marker compounds and performing method validation on the current HPLC method, to establish the linear range of the method and ensure reliable measurement of the marker peaks in future analyses. Further characterisation of the peaks of interest would enable confirmation of their identity, which would be significantly quicker using peak set B since optimizing the chromatography of fewer peaks would make validation easier. However given the low concentration of the peaks, it would take longer and be more costly to extract sufficient quantities for use as standards, and a pre-concentration step to the sample preparation method may be required.

It may be possible to select seasonal markers or identify temporal trends, which would aid interpretation in cases where there has been a delay between time of the crime and the collection of evidence and reference samples. Likewise, although the aim of the current study was to develop a method for exclusionary analysis and comparison, It may be possible, with further research, to identify individual or groups of peaks which are indicative of specific land uses or vegetation types. It is therefore recommended that further research should be conducted to identify land use markers for use in conjunction with existing soil databases to enable the new HPLC method to be used in intelligence cases.

5. Conclusions

This study has demonstrated the applicability of HPLC profiling as a highly discriminating tool for the comparison of soils taken from forensically relevant, close proximity locations within a single site, at four different sites. Nearly 100% of samples were correctly assigned to their location of origin at all four of the sites that were tested, using both sets of markers. This research provides improved methodology for sample analysis and two effective data analysis strategies to use in future geoforensic studies, allowing the data analysis method to be chosen to suit the priorities of the individual scenario. Where sample amounts are limited, the use of peak set A would be of greater value, as the peaks are taller and therefore sample concentration could be reduced, however where quicker analysis is required, peak set B offers more timely analysis. The results of this research show that this newly developed HPLC approach and associated data analysis methods provide significant scope for highly discriminatory, routine analyses to be performed on geoforensic samples from case relevant, close proximity locations, and could be applicable in a range of laboratories across the UK and internationally.

Acknowledgements

Thanks are due to Jasmine Ross (James Hutton Institute) and Katy Basford for their assistance in facilitating the sample analysis for this project.

This research was funded by the Engineering and Physical Sciences Research Council of the UK through their Centres for Doctoral Training programme, specifically the Security Science Doctoral Research Training Centre (UCLSECReT) based at University College London. Grant Title: Security Science Doctoral Training Centre — EP/G037264/1. Aspects of this work was funded by the Rural & Environment Science & Analytical Services Division of the Scottish Government.

References

[1] A. Ruffell, J. McKinley, Geoforensics Chichester, John Wiley & Sons Ltd, 2008.
[2] M. Stam, Soil as significant evidence in a sexual assault/attempted homicide case, in: K. Pye, D.J. Croft (Eds.), Forensic Geoscience: Principles, Techniques and Applications, Geological Society, London, 2004, pp. 295–299.
[3] Interpol, Forensic Examination of Soil Evidence, In 13th INTERPOL Forensic Science Symposium, Lyon, France, 2001 D1 175–D1 191.
[4] R. Sugita, S. Suzuki, Y.K, Forensic geology — a review: 2004-2007, In 15th INTERPOL Forensic Science Symposium, Lyon, France, 2007, pp. 81–85.
[5] R. Sugita, H. Yoshida, Forensic geology: review 2007 to 2009, In 16th International Forensic Science Symposium Interpol, Lyon, France, 2010.
[6] L.A. Dawson, S. Hillier, Measurement of soil characteristics for forensic applications, Surf. Interface Anal. 42 (5) (2010) 363–377.
[7] A. Ruffell, J. McKinley, Forensic geoscience: applications of geology, geomorphology and geophysics to criminal investigations, Earth Sci. Rev. 69 (March (3-4)) (2005) 235–247.
[8] R.M. Morgan, P.A. Bull, The use of grain size distribution analysis of sediments and soils in forensic enquiry, Sci. Justice 47 (3) (2007) 125–135.
[9] R.M. Morgan, P.A. Bull, Forensic geoscience and crime detection. Identification, interpretation and presentation in forensic geoscience, Minerva Medicolegale 127 (2) (2007) 73–89.
[10] R.M. Morgan, P.A. Bull, The philosophy, nature and practice of forensic sediment analysis, Prog. Phys. Geogr. 31 (1) (2007) 43–58.
[11] P.A. Bull, R.M. Morgan, J. Freudiger-Bonzon, A critique of the present use of some geochemical techniques in geoforensic analysis, Forensic Sci. Int. 178 (2–3) (2008) e35–e40.
[12] R.W. Fitzpatrick, Forensic comparison of soils, in: M. Tibbett, D.O. Carter (Eds.), Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Remains, CRC Press, Boca Raton, 2008, pp. 1–28.
[13] B. Minasny, A.B. McBratney, S. Salvador-Blanes, Quantitative models for pedogenesis — a review, Geoderma 144 (1–2) (2008) 140–157.
[14] D.W. Hopkins, The role of organisms in terrestrial decomposition, in: M. Tibbett, D.O. Carter (Eds.), Soil Analysis in Forensic Taphonomy: Chemical and Biological effects of Buried Human Remains, CRC Press, Boca Raton, 2008, pp. 53–66.
[15] L.A. Dawson, R.W. Mayes, Criminal and environmental soil forensics: soil as physical evidence in forensic investigations, in: B.L. Murphy, R.D. Morrison (Eds.), Introduction to environmental forensics, 3rd ed., Academic Press, Oxford, 2017, pp. 457–486.
[16] C.R. Bommarito, A.B. Sturdevant, D.W. Symanski, Analysis of forensic soil samples via high-performance liquid chromatography and ion chromatography, J. Forensic Sci. 52 (1) (2007) 24–30.
[17] J.A. Siegel, C. Precord, The analysis of soil samples by reverse phase high performance liquid chromatography using wavelength ratioing, J. Forensic Sci. 30 (2) (1985) 511–525.
[18] D.J. Reuland, W.A. Trinler, An investigation of the potential of high performance liquid chromatography for the comparison of soil samples, Forensic Sci. Int. 18 (1981) 201–208.
[19] D.J. Reuland, W.A. Trinler, M.D. Farmer, Comparison of soil samples by high performance liquid chromatography augmented by absorbance ratioing, Forensic Sci. Int. 52 (2) (1992) 131–142.
[20] G. McCulloch, P.A. Bull, R.M. Morgan, High performance liquid chromatography as a valuable tool for geoforensic soil analysis, Aust. J. Forensic Sci. (2016). http://dx.doi.org/10.1080/00450618.2016.1194474.
[21] K. Pye, Geological and Soil Evidence: Forensic Applications, CRC Press, Boca Raton, 2007.

[22] K. Simmons, United States Environmental Protection Agency. [Online], (2014) [cited 2016 April 15. Available from: https://www.epa.gov/quality/quality-system-and-technical-procedures-sesd-field-branches.

[23] R Core Team, R Foundation for Statistical Computing, Vienna, Austria. [Online]. Vienna, (2015) Available from: https://www.R-project.org/.

[24] J.O. Cerdeira, P.D. Silva, J. Cadima, M. Minhoto, Subselect: Selecting Variable Subsets. R package version 0.12-5. [Online]. Available from: https://CRAN.R-project.org/package=subselect.

[25] W.J. Krzanowski, Principles of Multivariate Analysis, Oxford university Press, Oxford, 1988.